

**Scientifically Based Research**

**Preliminary Review  
of  
Mathematics Curricular Materials**

**Idaho Department of Education  
June 2003**

Idaho Department of Education  
**Scientifically Based Research**

---

Susan Harrington  
Mathematics Coordinator  
Idaho Department of Education  
P.O. Box 83720  
Boise, ID 83720-0027  
Phone: 208-332-6932 • Fax: 208-334-4664  
Email: [sharring@sde.state.id.us](mailto:sharring@sde.state.id.us)

The team of teachers involved in the process of reviewing scientifically based research were convened by Dr. Dan Prinzing, Idaho Department of Education Curricular Materials Coordinator. They are representative of various K-12 grade levels and small/large Idaho school districts. Plans are being made for this team to re-convene next summer to continue this review process.

The following rationale regarding the process of reviewing scientifically based research was compiled by Frank Gallant, Ph.D.. At the time this research was completed, Dr. Gallant was Assistant Professor of Educational Leadership, University of Idaho Boise. Currently he is serving as interim superintendent of Basin School District #72, Idaho City, Idaho.

The results of this summer's "Round 1" scientifically based research preliminary review of mathematics curricular materials appear in Appendix III. Scores were based on the available information at the time of review. Programs not listed may or may not have sufficient evidence to qualify for inclusion into the scoring review process. All recommendations are contingent upon future federal guidelines.

---

# **Table of Contents**

<b>A Determination of Scientifically Based Research</b>	<b>1</b>
<b>Crosswalk between <u>Guidelines for judging the quality of a study</u> and the <u>Continuum of Evidence of Effectiveness Rubric</u></b>	<b>2</b>
<b>Appendix I <u>Guidelines for judging the quality of a study</u></b>	<b>7</b>
<b>Appendix II <u>Continuum of Evidence of Effectiveness Rubric</u></b>	<b>11</b>
<b>Appendix III Scientifically Based Research Preliminary Results</b>	<b>13</b>

## **A Determination of Scientifically Based Research**

The state of Idaho is undergoing a five year mathematics textbook adoption. Multiple teams of teachers worked to determine the acceptability of mathematics materials for adoption by the state. One aspect of this review has been to determine the scientifically based research (SBR) behind publishers' materials. One team of teachers with one professional researcher worked to determine the nature and the quality of the SBR as defined in the No Child Left Behind (NCLB) law. Based on guidance from the Idaho State Department of Education Bureau of Curriculum and Accountability and Bureau of Federal Programs, the team used a rubric (Continuum of Effectiveness: <http://www.ael.org/rel/csr/catalog/c.htm>) to gauge the quality of the scientifically based research (SBR) of the curricular materials which have been submitted to the Idaho State Department of Education. Additionally, based on guidance from the Office of Elementary and Secondary Education, U.S. Department of Education, guidelines for judging the quality of a study were presented in *Scientifically Based Research and the Comprehensive School Reform (CSR) Program*, August 2002, pp.7 – 10 (<http://www.ed.gov/offices/OESE/compreform/appendc.doc>). Thus based on the guidance of these two agencies, the Idaho team proceeded with its evaluations.

The rubric that the team used has three levels: Most Rigorous, Somewhat Rigorous, and Marginal. The committee rated the curricular materials by these three categories, and determined that a marginal rating was below the level of acceptability for Idaho's SBR. A marginal rating in no way implied that the quality of the curricular materials was unacceptable, or that the theory behind its development is weak. Rather the marginal rating means one of two things: either the curricular materials have not been submitted to a comprehensive analysis required by NCLB and the definition in the law for SBR, or the publisher failed to provide evidence of this rigorous analysis. The Idaho State Department of education will continue to evaluate mathematics materials over the next two years as part of this adoption process. The rubric itself provided structure to the analysis and a methodology for the team, and helped provide objective ratings for all the materials.

The Idaho team went one step farther in an effort to help substantiate its work. The team developed a crosswalk between the Federal Department of Education's guidelines (Appendix I), and the rubric (Appendix II). The crosswalk contains only the statement from the "Most Rigorous" category of the rubric. Different levels of the rubric which are less rigorous may more aptly apply to a specific instance from a publisher's material; however as a rule, the most rigorous classification provides a general description of the somewhat rigorous and marginal categories. That is, different levels of the rubric might be more applicable than the "Most Rigorous" category in specific instances. The Idaho SBR team assigned a rating of 3 to most rigorous, 2 to somewhat rigorous, and a 1 to marginally rigorous as it reviewed the publishers' materials. Some of the categories from the Federal Department of Education's (DOE) guidelines are redundant when compared with the rubric. This redundancy occurs because the DOE is attempting to be thorough and to provide multiple examples of its criteria, and the examples are themselves very much alike. The Idaho SBR team did not attempt to eliminate this redundancy, but rather responded to each individual category. Thus the reader may find some redundancy in the crosswalk. The redundancy was kept because the Idaho SBR team also tried to be thorough in its application of these criteria.

Furthermore, the rubric is more detailed than the DOE criteria; and as a result, the rubric has categories that are not addressed in the DOE criteria.

Thus the crosswalk is:

**Guidelines for judging the quality of a study**

**Rubric**

Criteria 1	Systematic and Empirical	
a.	Does the research have a sound theoretical foundation?	Does the model explain the theory behind its design, including references to the scientific literature, which elucidate why the model improves student achievement?
b.	Were the data obtained using observation or experimentation?	Have student achievement gains been shown using experimental and control groups created through large-scale random assignment or carefully matched comparison groups?
c.	Were the data collected from <i>all</i> appropriate groups of respondents and not just from certain groups?	Has the model been implemented in schools with characteristics similar to the target school: same grade levels, similar size, similar poverty levels, similar student demographics such as racial, ethnic, and language minority composition? <sup>1</sup>
d.	Were the data observed or collected from multiple subjects (teachers, students, schools, etc.)?	Has the model been fully implemented in multiple sites for more than 3 years?
e.	Are the research findings supported by measurable evidence?	Have the student achievement gains been sustained for three or more years?

**Guidelines for judging the quality of a study**

**Rubric**

Criteria 2	Rigorous data analysis	
a.	Does the research test the stated hypothesis, and do the findings justify the general conclusions drawn?	Have the student achievement gains been confirmed through independent, third-party evaluation?

<sup>1</sup> The team interpreted “Has the model been implemented?” to mean “were data available from schools?”

b.	Does the research report the sample size and the statistical procedures used?	Have the student achievement gains been confirmed through independent, third-party evaluation?
c.	Do the researchers analyze the data in a manner appropriate to the research question of interest? Are the statistical procedures used adequate for answering the research question?	Have the student achievement gains been confirmed through independent, third-party evaluation?
d.	Do the analysis methods correspond to the structure of the data? Does the analysis account for the complexities of the data, for missing data, for unique groupings, for changes in the data over time?	Have the student achievement gains been confirmed through independent, third-party evaluation?

**Guidelines for judging the quality of a study**

**Rubric**

Criteria 3	Reliable and valid data collection	
a.	Was data collection conducted professionally and consistently? For example, was there some system to ensure that different data collectors had the same focus and attention to detail?	Have student achievement gains been shown using experimental and control groups created through large-scale random assignment or carefully matched comparison groups?
b.	Were research biases minimized? Developers of reform models supply a natural example: was the evaluation of the reform model conducted by the model developers or by a third-party, independent evaluator?	Have the student achievement gains been confirmed through independent, third-party evaluation?
c.	Does the study look at the appropriate information to address its questions? Are the measures valid? That is, do the measures discussed and analyzed correspond to the concepts being studied?	Have the student achievement gains been confirmed through independent, third-party evaluation?
d.	Are the data reliable? Did repeated	Have the student achievement gains

	measurements on subjects taken under similar circumstances produce similar results? Do the data represent counts of actions, records, responses, etc., that directly reflect what the practice or program is supposed to be doing and affecting?	been confirmed through independent, third-party evaluation?
--	--	---

### **Guidelines for judging the quality of a study**

### **Rubric**

Criteria 4	Strong research design	
	Does the study follow an experimental or quasi-experimental design?	Have student achievement gains been shown using experimental and control groups created through large-scale random assignment or carefully matched comparison groups?
	Does the study design contain appropriate controls in order to be able to evaluate the effects of the condition of interest? Were the subjects of the research randomly assigned, or were there other within-condition or across-condition controls as part of the design?	Have student achievement gains been shown using experimental and control groups created through large-scale random assignment or carefully matched comparison groups?
	If subjects are not divided into the groups randomly, are the groups selected to ensure that subjects share similar background characteristics such as economic well-being or previous academic achievement? If not, does the study explain how statistical controls were used to account for these differences in background characteristics of the students in the study?	Has the model been implemented in schools with characteristics similar to the target school: same grade levels, similar size, similar poverty levels, similar student demographics such as racial, ethnic, and language minority composition?
	Did the research minimize alternative explanations for observed effects?	Has the model been implemented in schools with characteristics similar to the target school: same grade levels, similar size, similar poverty levels, similar student demographics such as racial, ethnic, and language minority composition? <sup>2</sup>

<sup>2</sup> The team interpreted “Has the model been implemented?” to mean “were data available from schools?”

	Does the study make a determination that the practice or program was used appropriately and fully as intended?	Has the model been fully implemented in multiple sites for more than 3 years?
--	--	---

**Guidelines for judging the quality of a study**

**Rubric**

Criteria 5	Detailed results that allow for replication	
	Are the findings clearly described and reported, free from technical terms and jargon?	
	Are the description of the design and the results of the research sufficiently detailed so that replication of the design is possible?	Has the model been replicated successfully in a wide range of schools and districts, e.g. urban, rural, suburban?
	Are the findings presented fairly and objectively?	Have student achievement gains been shown using experimental and control groups created through large-scale random assignment or carefully matched comparison groups?
	Are technical aspects of the study, such as statistical significance or confidence intervals made available and explained? Do the reports supply any supporting technical materials, perhaps in appendices?	Have student achievement gains been shown using experimental and control groups created through large-scale random assignment or carefully matched comparison groups?
	Is the presentation balanced? That is, are shortcomings reported as well as strengths? Were possible explanations provided for findings that ran counter to the researcher's expectations?	Have the student achievement gains been confirmed through independent, third-party evaluation?

**Guidelines for judging the quality of a study**

**Rubric**

Criteria 6	Expert Scrutiny	
	Has the research been accepted and published by a competitive, peer-reviewed scientific journal, or was it reported only in media such as newspapers, magazines, or trade journals?	Have the student achievement gains been confirmed through independent, third-party evaluation?



	<p>If the work was not published, is there evidence that it was reviewed by independent experts and subjected to external verification? If so, did the reviewers approve the study methodology and interpretation of the findings?</p>	<p>Have the student achievement gains been confirmed through independent, third-party evaluation?</p>
--	--	---

## Appendix I

### Guidelines for judging the quality of a study

The criteria for judging the quality of research studies are contained in the definition of scientifically based research in section 9101(37) of the reauthorized Elementary and Secondary Education Act (ESEA). Although there is no universally accepted standard, for the purposes of this publication, a **high quality study** meets *all* of the criteria described below. A **reasonable quality study** meets *all but one* of the criteria. For example, a reasonable quality study might be *systematic, empirical*, and use *rigorous data analysis* on *reliable and valid data*, but might use a longitudinal study *design* that does not involve random assignment to study groups or statistical controls on background characteristics.

#### Criteria 1: Systematic and empirical

High quality research is carried out in a manner that is consistent, disciplined, and methodical—not sloppy or haphazard. Such research shows evidence of careful planning and keen attention to detail. Empirical research is grounded in data drawn from observation or experiment; the claims being made are supported by measurable evidence, not opinion or speculation.

When evaluating research, consider the following:

Does the research have a sound theoretical foundation?

(See Research Consumer Questions, pp. 4 and 5.)

Were the data obtained using observation or experimentation?

Were the data collected from *all* appropriate groups of respondents and not just from certain groups? For example, does a school reform program that claims to benefit all students include special education students in its research? If the research uses test results for a given school, did all of the students in the school take the test?

Were the data observed or collected from multiple subjects (teachers, students, schools, etc.)?

Are the research findings supported by measurable evidence?

#### Criteria 2: Rigorous data analysis

Even the highest quality data are of little value unless analyzed thoughtfully and carefully. The definition of scientifically based research requires that data collected must be analyzed using methods that are appropriate for the task, and adequate to test the stated hypotheses and justify the general conclusions drawn. Failure to apply appropriate

methods could produce inaccurate or misleading findings.

Some key questions to consider about the data analysis include the following:

Does the research test the stated hypothesis, and do the findings justify the general conclusions drawn?

Does the research report the sample size and the statistical procedures used?

Do the researchers analyze the data in a manner appropriate to the research question of interest? Are the statistical procedures used adequate for answering the research question?

Do the analysis methods correspond to the structure of the data? Does the analysis account for the complexities of the data? for missing data? for unique groupings? for changes in the data over time?

For example, in school research studies that unfold over time, subjects may drop out of the study (for example, by moving out of a study school). Adequate data analyses address these issues.

### Criteria 3: Reliable and valid data collection

High quality data produce accurate and credible findings. Scientifically based research relies on measurements or observational methods that provide reliable and valid data across evaluators and observers and across multiple measurements and observations. Reliability implies that repeated measurements on subjects taken under similar circumstances or over time will produce similar results. If unreliable, the data may hinder the researcher's ability to discern real differences among subjects or programs. To be considered valid, the data collected must measure the outcomes they were designed to measure, (e.g. that student math knowledge is what is being measured, not students' ability to guess test answers). There must be a match between the research question and the observed behavior on which the research findings are based.

Questions about the quality of data collection include the following:

Was data collection conducted professionally and consistently? For example, was there some system to ensure that different data collectors had the same focus and attention to detail?

(e.g. training before data collection or interrater reliability tests)

Were research biases minimized? Developers of reform models supply a natural example: was the evaluation of the reform model conducted by the model developers or by a third-party, independent evaluator?

Does the study look at the appropriate information to address its questions? Are the measures valid? That is, do the measures discussed and analyzed correspond to the concepts being studied?

Are the data reliable? Did repeated measurements on subjects taken under similar circumstances produce similar results? Do the data represent counts of actions, records, responses, etc., that directly reflect what the practice or program is supposed to be doing and affecting?

#### Criteria 4: Strong research design

Studies must be designed to optimize the investigator's ability to answer the research question or hypothesis.

The following questions are relevant to research design:

Does the study follow an experimental or quasi-experimental design? That is, are the subjects in the study divided randomly into at least two groups, with at least one group using the practice or program of interest and one group not using it?

Does the study design contain appropriate controls in order to be able to evaluate the effects of the condition of interest? Were the subjects of the research randomly assigned, or were there other within-condition or across-condition controls as part of the design? (Random assignment of students is a way to ensure that it is the practice or program and not particular student characteristics that are producing the measured results.)

If subjects are not divided into the groups randomly, are the groups selected to ensure that subjects share similar background characteristics such as economic well-being or previous academic achievement? If not, does the study explain how statistical controls were used to account for these differences in background characteristics of the students in the study? (See criteria 2.)

Did the research minimize alternative explanations for observed effects?

Does the study make a determination that the practice or program was used appropriately and fully as intended?

#### Criteria 5: Detailed results that allow for replication

The results of high quality studies are presented in sufficient detail to allow for their replication, or to at least provide opportunities to build systematically on their findings. To increase their usefulness to practitioners, research findings must be reported in a way that makes them easily accessible and understood. The informed lay reader should be able to understand the study's design, methods, and findings.

When evaluating the quality of research reporting, consider the following:

Are the findings clearly described and reported, free from technical terms and

jargon?

Are the description of the design and the results of the research sufficiently detailed so that replication of the design is possible? For example, do researchers report the sample size (number of people or schools involved) and the statistical procedures used?

Are the findings presented fairly and objectively?

Are technical aspects of the study, such as statistical significance or confidence intervals made available and explained? Do the reports supply any supporting technical materials, perhaps in appendices?

Is the presentation balanced? That is, are shortcomings reported as well as strengths? Were possible explanations provided for findings that ran counter to the researcher's expectations?

#### Criteria 6: Expert Scrutiny

A strong study should be able to meet criticism by independent, expert reviewers. Peer reviewers, either from scientific journals or from an independent panel of experts in a given field, provide quality control in the form of a rigorous, objective, and scientific review of research. Research consumers can place more confidence in findings that have been subjected to expert review.

When evaluating research, consider the following:

Has the research been accepted and published by a competitive, peer-reviewed scientific journal, or was it reported only in media such as newspapers, magazines, or trade journals?

If the work was not published, is there evidence that it was reviewed by independent experts and subjected to external verification? If so, did the reviewers approve the study methodology and interpretation of the findings?

## Appendix II

### Continuum of Evidence of Effectiveness

	Most Rigorous	Somewhat Rigorous	Marginal
Theory/ Research Foundation	Does the model explain the theory behind its design, including references to the scientific literature, that elucidate why the model improves student achievement?	Does the model state the theory behind its design explaining how the model's components reinforce one another to improve student achievement?	Does the model explain the theory behind its design?
Evaluation- based Evidence of Effectiveness	Have student achievement gains been shown using experimental and control groups created through large-scale random assignment or carefully matched comparison groups?	Have student achievement gains been shown using between or within-school comparisons?	Have student achievement gains been shown for a single school?
	Has the model produced educationally significant pre and post intervention student achievement gains as reliably measured using appropriate assessments?	Has the model produced student achievement gains relative to district means or other comparison groups using appropriate assessment instruments?	Has the model produced improvements on other indicators of student performance , e.g. student attendance, graduation rates, or student engagement?
	Have the student achievement gains been sustained for three or more years?	Have the student achievement gains been sustained for one or two years?	Have other indicators of improved student performance been sustained for one or two years?
	Have the student achievement gains been confirmed through independent, third-party evaluation?	Has the model been evaluated by a state, district, or school evaluation team?	Has the model been evaluated by its developers?
Implemen- tation	Has the model been fully implemented in multiple sites for more than 3 years?	Has the model been fully implemented in the original site(s) for more than three years?	Has the model been fully implemented in the original pilot site(s) for a minimum of one school year?

	Has the model been implemented in schools with characteristics similar to the target school: same grade levels, similar size, similar poverty levels, similar student demographics such as racial, ethnic, and language minority composition?	Has the model been successfully implemented in at least one school with characteristics similar to the target school?	Is information on grade level, size, student demographics, poverty level, and racial, ethnic and language minority concentration available for the schools where the model has been implemented?
Replicability	Has the model been replicated successfully in a wide range of schools and districts, e.g. urban, rural, suburban?	Has the model been replicated in a number of schools or districts representing diverse settings?	Is full replication of the model being initiated in several schools?
	Have the replication sites have been evaluated, demonstrating significant student achievement gains comparable to those achieved in the pilot site(s)?	Have some replication sites been evaluated, demonstrating positive gains in student achievement?	Are promising initial results available from the replication sites?

## Appendix III

### Scientifically Based Research Preliminary Review June 2003 -- Round 1

#### Continuum of Evidence of Effectiveness

##### Most Rigorous:

Publisher	Title	Copyright	Grade Level
SRA McGraw Hill	Everyday Math	2004	K-6
Prentice Hall	Connected Mathematics	2004	6-8

##### Somewhat Rigorous:

Publisher	Title	Copyright	Grade Level
Harcourt School	Harcourt Math	2004	K-6
Houghton Mifflin	Houghton Mifflin Math	2005	K-6
Saxon	Saxon Math (K, 1, 2, 3, 54, 65, 76, 87, Alg.1/2)	2004	K-8
McDougal Littell	Middle Grades MathThematics (Books 1, 2, 3)	2002	6-8
McDougal Littell	Middle School Math (Courses 1,2,3,4 Passport to Alg/Geo)	2004	6-12
McDougal Littell	Algebra I, Geometry, Algebra II	2004	8-12
Renaissance Learning, Inc.	Accelerated Math	2002	9-12

Per No Child Left Behind's (2001) definition of "Scientifically Based Research" and the US Dept. of Education Comprehensive School Reform Guidelines (August 2002), The "Continuum of Evidence of Effectiveness" consists of a rubric with three-point scale and four criterion categories. The categories are as follows: Theory/Research Foundation, Evaluation-based Evidence of Effectiveness, Implementation, Replicability. Scores were based on the available information at the time of review. Programs not listed may or may not have sufficient evidence to qualify for inclusion into the scoring review process. All recommendations are contingent upon future federal guidelines.